WHAT IS CLAIMED:

1. A method for generating a compact representation of a first object, comprising:

(a) identifying a set of features corresponding to the first object;

(b) generating for each feature a hashing vector having n coordinates;

(c) summing the hashing vectors to obtain a summed vector; and

(d) creating an n*x-bit representation of the summed vector by calculating an x-bit value for each coordinate of the summed vector.

2. The method of claim 1, wherein the set of features is a vector.

3. The method of claim 1, wherein generating for each feature a hashing vector comprises:

determining a weight associated with each feature;

generating for each feature a hashing vector having n coordinates; and

multiplying each hashing vector by the weight determined for the corresponding feature.

4. The method of claim 1, wherein the object is a document.

5. The method of claim 4, wherein each feature is a word within the document.

6. The method of claim 1, wherein the object is a summary of another object.

7. The method of claim 1, wherein x is equal to 1.

8. The method of claim 1, wherein n is equal to 64.

9. The method of claim 1, further comprising:

repeating acts (a) – (d) for a second object to create a second n*x-bit representation; and

comparing the first and second n*x-bit representations to determine whether the first and second objects are similar.

10. The method of claim 9, further comprising discarding either one of the first or second objects.

11. The method of claim 1, further comprising:

repeating acts (a) – (d) for m objects to create m n*x-bit representations; and

grouping the m objects based on their corresponding n*x-bit representations.

12. The method of claim 11, further comprising compressing the objects by group.

13. The method of claim 1, wherein act (b) comprises generating for each feature a hashing vector having n coordinates, such that the hashing vectors are similar for similar features.

14. A method for generating a compact representation of an object, comprising:

generating a vector corresponding to the object, each coordinate of the vector being associated with a corresponding weight;

multiplying the weight associated with each coordinate in the vector by a corresponding hashing vector to generate a product vector;

summing the product vectors to obtain a summed product vector; and

generating a compact representation of the object using the summed product vectors.

15. The method of claim 14, wherein the weights are real numbers.

16. The method of claim 15, wherein the weights include values between zero and one.

17. The method of claim 14, wherein the object is a web document.

18. The method of claim 17, wherein the coordinates in the vector correspond to words in the web document.

19. The method of claim 18, further comprising:

assigning the weights for each coordinate of the vector as the number of occurrences of the word within the web document divided by the number of web documents contained in a collection of web documents that contain the word.

20. The method of claim 14, wherein values in the hashing vectors are generated using a pseudo random number generator seeded based on the coordinate corresponding to the hashing vector.

21. The method of claim 14, wherein each bit is generated based on the sign of the value of the coordinate.

22. A method comprising:

creating a similarity sketch for each of first and second objects based on an application of a hashing function to a vector representation of the first and second objects;

comparing, on a bit-by-bit basis, the similarity sketches for the first and second objects; and

generating a value defining the similarity between the first and second objects based on a correspondence in the bit-by-bit comparison.

23. The method of claim 22, further comprising:

determining that the first and second objects are similar when the value defining the similarity is greater than a predetermined threshold.

24. The method of claim 22, wherein creating the similarity sketch for each of the first and second objects further comprises:

generating a vector corresponding to the first and second objects, each coordinate of the vector being associated with a corresponding weight,

multiplying the weight associated with each coordinate in the vector by a corresponding hashing vector to generate a product vector,

summing the product vectors, and

calculating a bit corresponding to each coordinate of the summed product vector.

25. The method of claim 24, further comprising concatenating the generated bits.

26. A network device comprising:

at least one processor;

a database comprising a plurality of documents; and

a memory operatively coupled to the processor, the memory storing program instructions that when executed by the processor, cause the processor to remove similar objects from the database by comparing similarity sketches of pairs of objects in the database and removing one of the objects of the pair when the comparison indicates that the pair of objects are more similar than a threshold level of similarity, the processor generating the similarity sketches for each of the pair of objects based on application of a hashing function to vector representations of the objects.

27. A system for generating a compact representation of an object, comprising:

means for generating a vector corresponding to the object, each coordinate of the vector being associated with a corresponding weight;

means for multiplying the weight associated with each coordinate in the vector by a corresponding hashing vector to generate a product vector; and

means for summing the product vectors to obtain a summed product vector.

28. A computer-readable medium storing instructions for causing at least one processor to perform a method that generates a compact representation of an object, the method comprising:

generating a vector corresponding to the object, each coordinate of the vector being associated with a corresponding weight;

multiplying the weight associated with each coordinate in the vector by a corresponding hashing vector to generate a product vector;

summing the product vectors; and

generating the compact representation of the object using the summed product vector.

29. A method for generating a compact representation of an object, comprising:

generating an object vector corresponding to the object;

generating a hashing vector corresponding to each coordinate of the object vector;

summing the hashing vectors to obtain a summed vector;

calculating at least one bit corresponding to each coordinate of the summed product vector; and

generating a compact representation of the object by concatenating the calculated bits.